Cancer Exome Variant Intersections with 1KG Coding Variants

Lucas Lochovsky

Gerstein Lab

Genome Nettechotation Meeting

2012-08-14

Cancer Dataset Stats

Cancer	Paper Author	WGS/Exome	Somatic/Germline	Genome Build
Prostate Cancer	Berger et al.	WGS	Somatic	hg18/NCBI36
Malignant Melanoma	Pleasance et al.	WGS	Somatic	hg18/NCBI36
Multiple Myeloma	Chapman et al.	WGS	Somatic	hg18/NCBI36
Multiple Myeloma	Chapman et al.	Exome	Somatic	hg18/NCBI36

Cancer	Paper Author	# of Samples	# of Variants	# coding variants (from 1KG int)	Exome fraction (from 1KG int)
Prostate Cancer	Berger et al.	7	28,617	212	0.007408184
Malignant Melanoma	Pleasance et al.	1	33,344	293	0.008787188
Multiple Myeloma	Chapman et al.	22	1177	721	0.612574342
Multiple Myeloma	Chapman et al.	15	420	403	0.95952381

Review: TCGA Datasets

Cancer	WGS/Exome	Somatic/Germline	Center	Sequencer	# of Samples	# of Variants	# of Exome Variants	Exome Fraction	Genome Build
Colon adenocarcinoma (COAD)	Exome	Somatic	BCM	Illumina	52	22,147	21,920	99.0%	hg18/NCBI36
Colon adenocarcinoma (COAD)	Exome	Somatic	BCM	SOLiD	53	9,197	9,145	99.4%	hg18/NCBI36
Ovarian Serous Cystadenocarcinoma (OV)	Exome	Somatic	BCM	SOLiD	91	2,456	2,392	97.4%	hg18/NCBI36
Ovarian Serous Cystadenocarcinoma (OV)	Exome	Somatic	WUSTL	Illumina	88	6,192	5,613	90.6%	hg18/NCBI36
Rectum adenocarcinoma (READ)	Exome	Somatic	BCM	Illumina	12	1,716	1,703	99.2%	hg18/NCBI36
Rectum adenocarcinoma (READ)	Exome	Somatic	BCM	SOLiD	35	8,768	8,735	99.6%	hg18/NCBI36
Melanoma (Yale)	Exome	Somatic	Yale	Illumina?	89	25,489	-	100%	hg19/NCBI37

Cancer Variant Intersection with GENCODE Annotations

	Genes (Exons)	Pseudogenes	RNA	TFBSes
Berger Prostate	212	122	42	3503
Pleasance Melanoma	293	120	28	3969
Chapman Myeloma (WGS)	721	5	0	215
Chapman Myeloma (WES)	403	3	0	139

TCGA Cancer Exome Variants vs. 1KG Coding Variants

Somatic Variants - Actual Data

Cancer	# of cancer exome variants	# of 1KG coding variants	Intersection	Percentage of cancer variants in intersection	Percentage of 1KG coding variants in intersection
COAD BCM Illumina	21,452	514,269	1,700	7.92%	0.331%
COAD BCM SOLID	9119	514,269	702	7.70%	0.137%
OV BCM SOLID	2390	514,269	133	5.56%	0.0259%
OV WUSTL Illumina	5585	514,269	285	5.10%	0.0554%
READ BCM Illumina	1677	514,269	157	9.36%	0.0305%
READ BCM SOLID	8699	514,269	600	6.90%	0.117%

Somatic Variants - Average of 100 Runs Using Randomized Cancer Variant Positions

Cancer	# of cancer exome variants	# of 1KG coding variants	Intersection	Percentage of cancer variants in intersection	Percentage of 1KG coding variants in intersection
COAD BCM Illumina	21,452	514,269	607.65	2.83%	0.118%
COAD BCM SOLID	9119	514,269	254.97	2.80%	0.0496%
OV BCM SOLID	2390	514,269	67.28	2.82%	0.0131%
OV WUSTL Illumina	5585	514,269	167.76	3.00%	0.0326%
READ BCM Illumina	1677	514,269	46.90	2.80%	0.00912%
READ BCM SOLID	8699	514,269	245.17	2.82%	0.0477%

Somatic Variants - Actual:Random Percentage Ratio

Cancer	Ratio
COAD BCM Illumina	2.80
COAD BCM SOLID	2.75
OV BCM SOLID	1.98
OV WUSTL Illumina	1.70
READ BCM Illumina	3.35
READ BCM SOLID	2.45

Prostate cancer (Berger paper) vs. 1KG Coding Variants

Actual Data

# of cancer exome variants	# of 1KG coding variants	Intersection	Percentage of cancer variants in intersection	Percentage of 1KG coding variants in intersection
212	514,269	14	6.6%	0.00272%

Average of 100 Runs Using Randomized Cancer Variant Positions

# of cancer exome variants	# of 1KG coding variants	Intersection	Percentage of cancer variants in intersection	Percentage of 1KG coding variants in intersection
212	514,269	6.32	2.98%	0.00123%

Actual:Random
Percentage Ratio

2.22

Multiple Myeloma WGS (Chapman paper) vs. 1KG Coding Variants

Actual Data

# of cancer exome variants	# of 1KG coding variants	Intersection	Percentage of cancer variants in intersection	Percentage of 1KG coding variants in intersection
721	514,269	32	4.44%	0.00622%

Average of 100 Runs Using Randomized Cancer Variant Positions

# of cancer exome variants	# of 1KG coding variants	Intersection	Percentage of cancer variants in intersection	Percentage of 1KG coding variants in intersection
721	514,269	19.53	2.71%	0.00380%

Actual:Random Percentage Ratio

1.64

Multiple Myeloma WES (Chapman paper) vs. 1KG Coding Variants

Actual Data

# of cancer exome variants	# of 1KG coding variants	Intersection	Percentage of cancer variants in intersection	Percentage of 1KG coding variants in intersection
403	514,269	24	5.96%	0.00467%

Average of 100 Runs Using Randomized Cancer Variant Positions

# of cancer exome variants	# of 1KG coding variants	Intersection	Percentage of cancer variants in intersection	Percentage of 1KG coding variants in intersection
403	514,269	10.94	2.71%	0.00213%

Actual:Random Percentage Ratio

2.19

Malignant Melanoma (Pleasance paper) vs. 1KG Coding Variants

Actual Data

# of cancer exome variants	# of 1KG coding variants	Intersection	Percentage of cancer variants in intersection	Percentage of 1KG coding variants in intersection
293	514,269	13	4.44%	0.00253%

Average of 100 Runs Using Randomized Cancer Variant Positions

# of cancer exome variants	# of 1KG coding variants	Intersection	Percentage of cancer variants in intersection	Percentage of 1KG coding variants in intersection
293	514,269	8.14	2.78%	0.00158%

Actual:Random Percentage Ratio

1.60

Melanoma (Halaban paper) vs. 1KG Coding Variants

Actual Data

# of cancer exome variants	# of 1KG coding variants	Intersection	Percentage of cancer variants in intersection	Percentage of 1KG coding variants in intersection
25,489	514,269	771	3.02%	0.150%

Average of 100 Runs Using Randomized Cancer Variant Positions

# of cancer exome variants	# of 1KG coding variants	Intersection	Percentage of cancer variants in intersection	Percentage of 1KG coding variants in intersection
25,489	514,269	700.66	2.75%	0.136%

Actual:Random Percentage Ratio

1.10